



Cornell University
ILR School

Cornell University ILR School
DigitalCommons@ILR

Articles and Chapters

ILR Collection

4-1980

Regional Inequality and Other Sources of Income Variation in Colombia

Gary S. Fields
Cornell University, gfsf2@cornell.edu

T. Paul Schultz
Yale University

Follow this and additional works at: <http://digitalcommons.ilr.cornell.edu/articles>



Part of the [Economics Commons](#), and the [Labor Relations Commons](#)

This Article is brought to you for free and open access by the ILR Collection at DigitalCommons@ILR. It has been accepted for inclusion in Articles and Chapters by an authorized administrator of DigitalCommons@ILR. For more information, please contact hlmdigital@cornell.edu.

Regional Inequality and Other Sources of Income Variation in Colombia

Abstract

[Excerpt] Regional inequality is of interest for a variety of reasons: planning development policies aimed at alleviating poverty and reducing personal inequality, gauging the degree of a country's labor market integration, understanding patterns of population movement in general and labor force migration in particular, predicting future urbanization, and characterizing the poor. Policymakers often aim development programs at particular target groups such as those living in certain regions of a country. In this paper we analyze the determinants of incomes and income inequality in one less developed country, Colombia, examining both personal and regional aspects. The results help clarify the potential of a development strategy emphasizing the poorest regions of the country.

Our specific objectives are to measure the relative importance of personal and regional effects on income variation in Colombia and to determine within relatively homogeneous segments of the labor force how place of residence is associated with personal income levels and dispersion. In Section II we explore income differences across a number of dimensions (education, sex, age, type of employment, and region) with particular reference to regional inequality. We then turn, in Section III, to more formal procedures (analysis of variance and regression) for systematically analyzing the relationship between income and age, education, type of employment, and region, and for quantifying these effects. The paper concludes with some implications of the empirical findings.

Keywords

development, Colombia, income, inequality, migration

Disciplines

Economics | Labor Relations

Comments

Suggested Citation

Fields, G. S. & Schultz, T. P. (1980). Regional inequality and other sources of income variation in Colombia [Electronic version]. *Economic Development and Cultural Change*, 28(3), 447-467.

Required Publisher Statement

© The University of Chicago Press. Reprinted with permission. All rights reserved.

Regional Inequality and Other Sources of Income Variation in Colombia*

Gary S. Fields
Cornell University

T. Paul Schultz
Yale University

I. Introduction

Regional inequality is of interest for a variety of reasons: planning development policies aimed at alleviating poverty and reducing personal inequality, gauging the degree of a country's labor market integration, understanding patterns of population movement in general and labor force migration in particular, predicting future urbanization, and characterizing the poor. Policymakers often aim development programs at particular target groups such as those living in certain regions of a country.¹ In this paper we analyze the determinants of incomes and income inequality in one less developed country, Colombia, examining both personal and regional aspects.² The results help clarify the potential of a development strategy emphasizing the poorest regions of the country.

* This research is supported in part by AID contract otr-1432 and a grant from the Rockefeller-Ford Foundations' Research Program on Population and Development Policy, and is facilitated by Rockefeller Foundation grant RF-70051 to Yale's Economic Demography Program. An earlier version of this paper was presented at the Conference on Poverty and Development in Latin America, Yale University. We have benefited from the helpful comments of Philip Musgrove and Juan Buttari on the earlier draft. We wish to thank Ruth Daniel, Helena Jaramillo, and Judith Oder for their invaluable research assistance in preparing the data for this paper.

¹ In the development literature, the best-known studies of regional inequality are those of Simon Kuznets, "Quantitative Aspects of the Economic Growth of Nations. VIII. Distribution of Income by Size," *Economic Development and Cultural Change* 11, no. 2, pt. 2 (January 1963): 1-80; Jeffrey Williamson, "Regional Inequality and the Process of National Development: A Description of the Patterns," *ibid.*, 13, no. 4, pt. 2 (July 1965): 33-84.

² Noteworthy in the Colombian context is the research on regional income inequality by R. Albert Berry and Miguel Urrutia, *Income Distribution in Colombia* (New Haven, Conn.: Yale University Press, 1976); Richard R. Nelson, T. Paul Schultz, and Robert Slighton, *Structural Change in a Developing Country: Colombia's Problems and Prospects* (Princeton, N.J.: Princeton University Press, 1971); and Robert Prieto, *Estructura del gasto y distribución del ingreso familiar en cuatro ciudades Colombianas: 1967-68* (Bogotá: Centro de Estudios sobre Desarrollo Económico, Universidad de Los Andes, 1971).

Our specific objectives are to measure the relative importance of personal and regional effects on income variation in Colombia and to determine within relatively homogeneous segments of the labor force how place of residence is associated with personal income levels and dispersion. In Section II we explore income differences across a number of dimensions (education, sex, age, type of employment, and region) with particular reference to regional inequality. We then turn, in Section III, to more formal procedures (analysis of variance and regression) for systematically analyzing the relationship between income and age, education, type of employment, and region, and for quantifying these effects. The paper concludes with some implications of the empirical findings.

II. Income Variation in Colombia: Tabulations

The data for this paper are taken from the fourteenth Colombian Census of Population (October 1973). A 4% public use sample of 860,000 census returns was provided to us by the National Statistical Office (known by its Spanish acronym, DANE).

To determine income, the census asked: "What was your income in pesos last month?" Our concern in this paper is with the determinants of personal income and its correlates. Accordingly, children under the age of 10 and persons not in the labor force are not dealt with. After excluding nonincome recipients in the economically active population (e.g., unremunerated family workers) and nonrespondents, and also domestic servants who also receive much of their income in kind, we are left with a working sample of 141,841 employees and employers.

Table 1 presents average incomes by various characteristics of the individual. The gross differentials (without cross-classification) are of the following orders of magnitude: eleven-to-one ratio between persons with higher education and persons with none; four-to-one between prime age workers and the very young; a 25% differential advantage for men over women; four-to-one ratio between the richest department and the poorest; three-to-one between urban workers and rural workers; and 60% more for employers and the self-employed than for wage and salary employees.

It is sometimes thought that gross income differentials, like those observed in table 1, arise from failure to standardize for other factors determining income. Regional inequality in particular is attributed by some to differences in the educational composition of various departments' labor forces. Were this view correct, a weighted average of incomes within educational categories would be more nearly equal across departments than the department averages.

Cross-tabulated data are shown in tables 2 and 3. Table 2 presents for the 23 departments of Colombia the sample estimates of average monthly incomes of men and women by four educational classes: no schooling, some or all primary schooling (1-5), some or all secondary schooling (6-11), and some or all higher education (12+). Employers

and employees are treated here together. Beneath each entry in parentheses is the number of individuals on which the average income is based.

The average income differentials noted in table 1 remain in attenuated form in finer breakdowns. Income increases with education, not only in the country as a whole, but for men and women in every department. The same regularity exists in Venezuela,³ but in a few instances workers with no schooling receive higher incomes than those with some primary schooling, for example, in the Federal District of Caracas.

Another similarity between the two countries is that women's incomes are less than men's. Once again, this is true for each educational

TABLE 1
MEAN MONTHLY INCOMES (in Pesos) IN COLOMBIA, OCTOBER 1973,
FOR PERSONS WITH VARIOUS CHARACTERISTICS

Department of residence:*	Rural/urban:
Antioquia 1,536	Rural 536
Atlántico 1,872	Urban 1,676
Bogotá, D.E. 2,694	
Bolívar 1,324	Education:†
Boyacá 897	None 610
Caldas 1,253	Primary 982
Cauca 819	Secondary 2,337
César 1,391	University 6,898
Córdoba 1,039	
Cundinamarca 946	Sex:
Chocó 656	Male 1,540
Huila 1,093	Female 1,232
La Guajira 1,726	
Magdalena 1,279	Employment status:‡
Meta 1,360	Employer 2,093
Nariño 667	Employee 1,275
Norte de Santander 1,073	
Quindío 1,337	Age:
Risaralda 1,372	10-19 532
Santander 1,151	20-24 1,034
Sucre 1,121	25-29 1,523
Tolima 1,205	30-34 1,830
Valle 1,608	35-44 1,941
	45-54 1,961
	55-64 1,710
	65 and over 1,279

* Colombia is divided into 22 departments, analogous to states, and the special district of Bogotá. A number of frontier territories and small islands (less than 2% of the population) are excluded from the census sample.

† Each category under "Education" refers to individuals who either attended or completed that level. The number of years of each are primary (grades 1-5), secondary (grades 6-11), and higher (12+).

‡ We distinguished three types of income recipients: (1) "Employees" include day workers (*jornaleros*), wage laborers (*obrerros*), and salaried employees (*empleados*). (2) "Employers" include the self-employed (*trabajadores independientes*) and employers of others (*patrones*). (3) A residual category, excluded here, consists of domestic servants and unpaid family workers.

³ T. Paul Schultz, "Determinants of Internal Migration in Venezuela: An Application of the Polytomous Logistic Model" (paper presented at the Econometrics Society Third World Congress, Toronto, August 1975).

TABLE 2

MALE AND FEMALE MONTHLY INCOMES, OCTOBER 1973, FOR COLOMBIA BY DEPARTMENT AND EDUCATION (in Pesos)

DEPARTMENT	EDUCATION, MALE					EDUCATION, FEMALE					BOTH SEXES (Total)
	None	Primary	Secondary	Higher	All Males	None	Primary	Secondary	Higher	All Females	
Antioquia	703 (3,429)	1,092 (10,542)	2,732 (3,566)	7,997 (629)	1,580 (18,166)	468 (223)	769 (1,564)	1,755 (1,637)	3,386 (156)	1,315 (3,580)	1,536 (21,746)
Atlantico	820 (600)	1,358 (2,570)	2,710 (1,434)	7,104 (295)	2,034 (4,899)	572 (83)	842 (567)	1,588 (674)	2,801 (93)	1,310 (1,417)	1,872 (6,316)
Bogotá, D.E.	912 (815)	1,337 (8,744)	2,974 (5,910)	8,370 (2,081)	2,702 (17,550)	559 (323)	776 (2,818)	1,852 (3,205)	3,678 (583)	1,508 (6,929)	2,694 (21,479)
Bolívar	753 (978)	1,091 (1,611)	2,304 (647)	7,317 (86)	1,389 (3,322)	466 (127)	651 (320)	1,500 (278)	3,421 (28)	1,036 (753)	1,324 (4,075)
Boyacá	343 (910)	654 (2,920)	2,478 (484)	5,474 (89)	888 (4,403)	252 (92)	497 (253)	1,509 (237)	3,568 (20)	960 (602)	897 (5,005)
Caldas	712 (665)	961 (2,639)	2,447 (585)	7,754 (84)	1,282 (3,973)	390 (42)	608 (305)	1,514 (301)	3,116 (22)	1,084 (670)	1,253 (4,643)
Cauca	412 (573)	633 (1,564)	1,887 (236)	5,758 (50)	809 (2,423)	325 (83)	500 (194)	1,574 (143)	3,228 (8)	876 (428)	819 (2,851)
César	918 (422)	1,182 (767)	2,542 (196)	8,466 (28)	1,436 (1,413)	547 (48)	714 (98)	1,823 (76)	4,060 (4)	1,111 (226)	1,391 (1,639)
Córdoba	616 (1,268)	849 (1,446)	2,791 (294)	6,919 (43)	1,025 (3,051)	377 (112)	602 (164)	2,192 (152)	2,724 (7)	1,134 (435)	1,039 (3,486)
Chocó	248 (307)	714 (359)	1,795 (91)	6,621 (14)	763 (771)	89 (207)	253 (69)	1,550 (58)	4,180 (3)	410 (337)	656 (1,108)
Cundinamarca	518 (1,313)	760 (4,310)	2,098 (695)	5,175 (103)	926 (6,421)	481 (100)	635 (384)	1,662 (370)	2,524 (20)	1,095 (874)	946 (7,295)
Huila	750 (664)	867 (1,681)	2,464 (281)	6,066 (46)	1,095 (2,672)	477 (43)	731 (166)	1,404 (202)	3,450 (8)	1,081 (419)	1,093 (3,091)
La Guajira	912 (99)	1,592 (294)	2,798 (121)	7,800 (10)	1,860 (524)	486 (18)	781 (53)	1,732 (64)	3,500 (1)	1,209 (136)	1,726 (660)
Magdalena	710 (816)	1,100 (1,130)	2,518 (362)	6,362 (44)	1,281 (2,352)	445 (50)	841 (127)	1,651 (187)	3,896 (9)	1,268 (373)	1,279 (2,725)
Meta	750 (307)	1,070 (906)	2,730 (246)	6,863 (26)	1,380 (1,485)	655 (20)	874 (103)	1,563 (106)	4,100 (5)	1,236 (234)	1,360 (1,719)

TABLE 2 (Continued)

DEPARTMENT	EDUCATION, MALE					EDUCATION, FEMALE					BOTH SEXES (Total)
	None	Primary	Secondary	Higher	All Males	None	Primary	Secondary	Higher	All Females	
Nariño.....	390 (930)	513 (2,907)	1,937 (349)	5,606 (59)	674 (4,245)	206 (173)	370 (561)	1,548 (184)	3,434 (20)	636 (938)	667 (5,183)
Norte de Santander	523 (994)	881 (2,371)	2,390 (481)	6,373 (72)	1,076 (3,918)	316 (71)	603 (295)	1,591 (272)	2,915 (22)	1,056 (660)	1,073 (4,578)
Quindío.....	802 (284)	1,066 (1,216)	2,276 (370)	6,604 (49)	1,402 (1,919)	347 (33)	510 (162)	1,424 (140)	2,982 (18)	983 (353)	1,337 (2,272)
Risaralda.....	673 (396)	1,083 (1,840)	2,576 (505)	6,572 (66)	1,423 (2,807)	484 (30)	828 (329)	1,376 (230)	3,869 (25)	1,140 (614)	1,372 (3,421)
Santander.....	472 (1,580)	927 (4,206)	2,706 (891)	6,517 (159)	1,184 (6,836)	304 (180)	630 (643)	1,429 (585)	3,488 (41)	993 (1,449)	1,151 (8,285)
Sucre.....	569 (654)	1,076 (615)	3,109 (175)	5,228 (23)	1,158 (1,467)	481 (70)	588 (81)	1,449 (78)	3,243 (4)	890 (233)	1,121 (1,700)
Tolima.....	753 (1,143)	962 (2,894)	2,489 (579)	7,870 (86)	1,226 (4,702)	401 (73)	597 (247)	1,425 (360)	3,058 (16)	1,061 (696)	1,205 (5,398)
Valle.....	666 (1,755)	1,162 (8,698)	2,569 (3,012)	8,502 (540)	1,685 (14,005)	479 (216)	819 (1,530)	1,660 (1,387)	3,737 (139)	1,277 (3,272)	1,608 (17,277)
Total.....	634 (20,902)	1,027 (66,230)	2,670 (21,504)	7,806 (4,682)	1,546 (113,318)	398 (2,417)	713 (11,033)	1,681 (10,926)	3,504 (1,252)	1,232 (25,628)	1,488 (138,946)
Regional mean income.....	649	997	2,492	6,840	...	406	653	1,597	3,407
Variance.....	33,339	67,511	126,103	1,389,357	...	20,987	30,265	38,025	217,184
Coefficient of variation.....	.288	.253	.133	.151364	.244	.116	.133
Regional mean logarithm of income.....	6.43	6.87	7.81	8.82	...	5.97	6.45	7.37	8.12
Variance of logarithm of income.....	.115	.0708	.0197	.0233179	.0850	.0118	.0183

NOTE.—Number of individuals on which the average income is based is in parentheses.

TABLE 3
RURAL AND URBAN MEAN INCOMES OF MALES BY AGE AND EDUCATION

AGE	NONE		PRIMARY		SECONDARY		HIGHER	
	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Employees:								
10-19.....	400 (361)	412 (143)	377 (914)	513 (945)	632 (36)	835 (418)	...	1,156 (4)
20-24.....	485 (291)	634 (118)	492 (722)	820 (923)	923 (47)	1,375 (783)	...	3,054 (99)
25-29.....	459 (202)	673 (84)	561 (499)	1,020 (798)	1,322 (41)	2,046 (684)	8,500 (2)	5,236 (161)
30-34.....	517 (184)	685 (94)	565 (401)	1,181 (733)	1,195 (17)	2,595 (415)	3,000 (2)	6,590 (159)
35-44.....	494 (398)	886 (214)	763 (621)	1,343 (1,258)	1,019 (14)	3,309 (525)	...	9,211 (136)
45-54.....	483 (274)	767 (196)	615 (392)	1,397 (771)	2,464 (7)	3,766 (235)	8,000 (1)	9,551 (54)
55+.....	426 (268)	623 (160)	560 (252)	1,131 (377)	950 (4)	3,206 (76)	1,000 (1)	7,601 (28)
Total.....	463 (1,978)	688 (1,009)	543 (3,801)	1,053 (5,805)	1,060 (166)	2,158 (3,136)	5,333 (6)	6,520 (641)
Employers:								
10-19.....	361 (42)	404 (29)	545 (98)	701 (93)	550 (2)	1,044 (32)
20-24.....	594 (40)	507 (18)	505 (115)	1,086 (184)	1,005 (5)	2,813 (115)	...	4,417 (12)
25-29.....	820 (50)	2,061 (28)	706 (158)	1,653 (231)	1,683 (17)	2,972 (139)	...	8,338 (42)
30-34.....	665 (65)	888 (27)	730 (190)	1,616 (290)	2,308 (10)	3,680 (148)	...	8,425 (45)
35-44.....	798 (165)	1,047 (111)	957 (325)	1,938 (635)	2,348 (15)	4,777 (275)	...	11,276 (62)
45-54.....	832 (144)	1,462 (92)	900 (262)	2,281 (457)	1,519 (16)	5,698 (186)	9,175 (4)	11,427 (48)
55+.....	659 (191)	977 (93)	832 (239)	2,178 (372)	1,311 (9)	5,759 (128)	...	7,714 (36)
Total.....	718 (697)	1,116 (398)	798 (1,387)	1,856 (2,262)	1,745 (74)	4,326 (1,023)	9,175 (4)	9,419 (245)

NOTE.—Number of individuals on which the average income is based is in parentheses.

group in a given department or province as well as in a comparison of the aggregate means.

Colombia, like Venezuela, exhibits substantial variation in average incomes across regions. For Colombian males with no education, for example, the average income in the richest department (Bogotá) is more than three times higher than in the poorest department (Chocó).⁴ Wide interregional differences are observed in all educational categories for both sexes. Interregional income variation by education group follows a common pattern in the two countries: (1) The absolute variance of incomes increases with education attainment, but (2) the variance of the logarithms of income and the coefficient of variation, which measure relative inequality independently of the mean, decline in both countries as educational level increases.⁵ Relative variation in regional incomes is thus greater for the least educated, possibly because skilled labor markets are closer to equilibrium due to greater mobility of the highly educated.⁶

Another important factor influencing income is age. Table 3 reports mean incomes for the Colombian sample of men broken down by rural-urban residence, age, and education for employees and employers.⁷ The age income profiles peak in the cross section at middle age. The systematic positive relationship between income and education is found for all age groups. As anticipated, employer incomes are somewhat higher than employee incomes given age, education, and rural-urban locations. The added returns to being an employer appear to grow systematically with age even though a growing fraction are becoming employers.

Rural and urban income differences remain pronounced in table 3 even within these cells. It is interesting that the absolute differentials appear to increase with education up through the secondary level. Note too the virtual absence of persons with higher education in rural areas. This may be because higher education is only offered in the cities or because migration to the cities is selective of the most highly qualified rural persons.⁸ The increase in the rural-urban income differential with

⁴ Somewhat surprisingly, for males with university education incomes are higher in two departments (César and Valle) than in Bogotá. We cannot tell whether this is because of greater relative scarcity of highly educated workers in those departments or because of measurement error. The gross income ratio between Bogotá and Chocó is four to one. The gross differential is greater than the within-educational category differential because Chocó also has fewer highly educated people proportionately than does Bogotá.

⁵ Except among the higher educated in Colombia.

⁶ Aba Schwartz, "On Efficiency of Migration," *Journal of Human Resources* 6, no. 2 (April 1971): 193-205; and Schultz, "Determinants of Internal Migration in Venezuela."

⁷ The Venezuelan data did not include age tabulations, so intercountry comparisons on this dimension are not possible.

⁸ Simon Kuznets, "Introduction: Population Redistribution, Migration and Economic Growth," in Hope T. Eldridge and Dorothy S. Thomas, eds., *Population Redistribution and Economic Growth* (Philadelphia: American Philosophical Society, 1964); and David Turnham, *The Employment Problem in Less Developed Countries: A Review of Evidence*, Employment Series, no. 1 (Paris: OECD, 1971).

educational level and the lack of highly educated rural workers provide evidence that such a selective-migration process is going on in Colombia.

After standardizing for each of these variables singly and together in Colombia, income differences are still substantial within sex, education, employment status, and age cells across departments and particularly between rural and urban areas.⁹ These remaining regional income differences could arise from various sources, among them: (1) relevant characteristics of workers, such as their actual job experience or ability, are omitted; (2) money income differences could represent a form of compensating variation for regional price variations, income in kind, and amenity levels; (3) regional labor markets may be in disequilibrium, paying different real wages for similar services; and (4) errors in measurement in addition to purely random variability are impounded in the residual.

To summarize systematically the many comparisons of personal and regional income differentials, a statistical framework is useful. For this purpose, a linear model is adopted in Section III.

III. Income Variation in Colombia: Linear Model

This section presents the results of estimating a linear model of income determination in Colombia. Two closely related linear models are used: analysis of variance (ANOVA) and multiple regression.

Analysis of variance decomposes overall income variance (or the variance in the logarithm of income) into within-category and between-category components, measures the direct contribution of each set of categories to total variance, and tests the marginal statistical significance of these effects.¹⁰ In comparison with other decomposable measures of inequality, specifically the Theil index of inequality and the Gini coefficient, ANOVA has three advantages: (1) generally accepted tests of statistical significance are available for ANOVA and not for the other decomposition procedures;¹¹ (2) the log-variance measure of inequality

⁹ Another table, available from the authors upon request, presents a detailed cross-classification of the population by sex, education, age, and department sub-groupings (1,472 cells in all).

¹⁰ ANOVA procedures have long been used to analyze experimental or quasi-experimental data, but their application to economic problems is quite limited. See Ronald A. Fisher, *Statistical Methods for Research Workers*, 7th ed. (Edinburgh: Oliver & Boyd, 1938); and George W. Snedecor, *Calculation and Interpretation of Analysis of Variance* (Ames, Iowa: Collegiate Press, 1934). In particular, on the problem of determining income and income inequality, work is just beginning. See T. Paul Schultz, "The Distribution of Personal Income: Case Study of the Netherlands" (Ph.D. diss., M.I.T., 1965); Carlos G. Langoni, "Distribuição da renda e desenvolvimento econômico do Brasil," *Estudos econômicos* 2 (October 1972): 5-88, and "Income Distribution and Economic Development: Brazilian Case" (paper presented at Econometric Society Third World Congress, Toronto, August 1975); Albert Fishlow, "Brazilian Income Size Distribution—Another Look," *Dados* 2 (1973): 10-80; and Carmel U. Chiswick, "Income Distribution in Thailand: Application of the Theil Index to Income Inequality" (World Bank working paper, DRC Series B-2, Washington, D.C., July 1976).

¹¹ This advantage is less important in our work than in most other income-distribution research because of our exceptionally large sample. Given the very large

attaches greater importance to the relative income status of the poor;¹² and (3) because of ANOVA's equivalence to multiple regression, quantitative effects of regional and personal variables on income may be estimated.¹³

Below we present the results of ANOVA on a sample of male employees in Colombia.¹⁴ The working sample consists of every fifth individual in the 4% DANE census file.¹⁵ The sample size is 16,695. The dependent variable in the empirical research is the natural logarithm of monthly income in pesos.¹⁶ Persons without incomes and the unemployed are attributed 1 peso per month in order to include them in the log-variance calculation. The explanatory categories are education, age, and place of residence. Four educational categories are distinguished: none, primary (some or all), secondary (some or all), and higher (some or all). There are seven age categories: 10-19, 20-24, 25-29, 30-34, 35-44, 45-54, 55 and over. Two place-of-residence variables are analyzed. One is rural-urban. The other is department of residence at three different levels: the department itself (23 in number), groups of departments (11), and geographic regions (6). The geographic distinctions analyzed are shown in table 4.

Analysis of Variance: Main-Effects Model

A main-effects model without interactions is reported in table 5. The first column indicates the simple association between the logarithm of income

sample size, virtually any basis for grouping the data according to personal, demographic, economic, social, or geographic information would reduce the standard error of estimate sufficiently to satisfy the *F*-test of statistical significance.

¹² For a comparison of the various decomposition procedures and a review of empirical studies in less developed countries, see Gary S. Fields, "Decomposing LDC Inequality" (Center Discussion Paper no. 263, Economic Growth Center, Yale University, July 1977).

¹³ We utilize this similarity below to estimate the extent of interregional income inequality due to interdepartmental variation in education, age, and rural-urban population distribution.

¹⁴ Women are excluded because they are thought more likely than men to work part time, which complicates interpretations of income variability. Also, for men, age may be a reasonable proxy for labor force experience, which is why age is included as an explanatory variable; for women, actual labor market experience may vary substantially even within a given age group. For comparative purposes, all statistical exercises were also performed on a corresponding sample of male employees (which also included independent workers). These results were substantially similar to those reported here for employees but are omitted to save space. See Gary S. Fields and T. Paul Schultz, "The Specification of an Income Function for Colombia" (unpublished paper, Yale University, 1978).

¹⁵ The smaller sample was used both to save on computational costs and to remain within the storage capacity of the Yale computer on some of the more complicated runs.

¹⁶ This transformation of income seems advisable for several reasons. For one thing, the statistical tests applied to ANOVA resolutions of variance are based on the assumption that the dependent variable is normally distributed. In Colombia, the log of income is quite nearly normally distributed, more so than is income itself. Furthermore, the log variance of income, as an index of inequality, is more sensitive to inequality associated with low incomes of the poor than are the variance and most other inequality measures.

TABLE 4
GEOGRAPHIC DISTINCTIONS ANALYZED

Department	Groups of Departments	Region
1. Atlántico	A	I
2. Bolívar	B	I
3. Córdoba	B	I
4. Sucre	B	I
5. Magdalena	C	I
6. La Guajira	C	I
7. César	C	I
8. Antioquia	D	II
9. Caldas	D	II
10. Quindío	D	II
11. Risaralda	D	II
12. Valle	E	III
13. Chocó	E	III
14. Cauca	F	III
15. Nariño	F	III
16. Tolima	G	IV
17. Huila	G	IV
18. Meta	H	IV
19. Boyacá	I	V
20. Santander	I	V
21. Norte de Santander	I	V
22. Bogotá, D.E.	J	VI
23. Cundinamarca	K	VI

and each set of explanatory categories; it is comparable to the simple zero-order correlation in the two-category case. The remainder of table 5 presents five analyses of variance (ANOVA) based on various alternative geographic distinctions, also including age and education categories. By conventional statistical standards all of the main effects are highly significant at confidence levels in excess of .999.

The relative importance of the various effects may be interpreted in either of two ways. First, there is the proportion of the variance in the logarithms of income directly explained by each set of explanatory categories. Second, the marginal *F*-ratio is shown, which deflates the explained variance by the number of categories considered and formally expresses the resulting reduction in standard error of estimate as a ratio to that anticipated from a random set of categories in a normally distributed population. For employees, education provides the most information in predicting personal incomes, in the sense of explaining between 12% and 19% of log variance. Its statistical significance is also greatest with *F*'s in excess of 1,000. Next in importance according to the *F*-criterion is the one-way rural-urban distinction, accounting for from 1.6% to 3.1% of the log variance with an *F* of 400-800. Then, we have the seven age categories which account for 6% or 7% of the log variance in incomes and receive an *F* of around 300. Lowest in explanatory power is region (see table 5). The regional distinctions, though still highly significant by conventional standards with *F*'s between 50 and 90, explain less than might have been antici-

TABLE 5

ANALYSIS OF VARIANCE: MAIN EFFECTS, MALE EMPLOYEES

MAIN EFFECTS	GEOGRAPHIC DISTINCTION										
	ZERO- ORDER CORRE- LATION (Eta)	1. Rural-Urban		2. Regions		3. Groups of Departments		Departments			
		Propor- tion of Variance Explained	<i>F</i> -Ratio, Marginal	Propor- tion of Variance Explained	<i>F</i> -Ratio, Marginal	Propor- tion of Variance Explained	<i>F</i> -Ratio, Marginal	4. Without Rural-Urban		5. With Rural-Urban	
								Propor- tion of Variance Explained	<i>F</i> -Ratio, Marginal	Propor- tion of Variance Explained	<i>F</i> -Ratio, Marginal
Education level (4)48	.129	1,131	.194	1,656	.165	1,388	.164	1,388	.120	1,038
Age group (7)31	.064	282	.072	307	.071	298	.071	299	.064	278
Rural-urban (2)37	.031	805016	404
Regions (6)21014	70
Groups of departments (11)037	92
Departments (23)32043	49	.028	33
Covariance115045073074140	...
Total explained339	893	.326	595	.345	458	.352	288	.367	299

NOTE.—Logarithm of income: mean = 6.52, variance = 1.54; sample size = 16,542. The number of explanatory categories is in parentheses. All effects statistically significant at .001 level.

Economic Development and Cultural Change

pated given the prominence accorded interregional variation in studies of income distribution in Colombia and elsewhere. The six regions account for only 1.4% of the log variance; the 11 labor market groupings of departments account for 3.7%, and the full 23 departments explain 4.3%. These values are much lower than the proportion of variance explained by education and age. In terms of the *F*-test, the 11-way grouping of departments appears the most significant. In all, slightly more than one-third of the variance of the logarithm of income is explained by these three or four sets of categories.

Exploring covariation among the explanatory variables, we find that the direct effect of age is not greatly influenced by the inclusion of various geographic distinctions, varying narrowly from 6.4% to 7.2% of the explained variance. Education, however, differs between rural and urban areas more than it does by department or region. When the rural-urban distinction is considered (ANOVA 1) the direct effect of education is 12.9%, but education's effect rises to 19.4% when only the six regions are included (ANOVA 2). On the other hand, the covariance effect falls from 11.5% to 4.5%, confirming the strong association between education, age, and the rural-urban categorization. Once the rural-urban distinction has been included, it is clear from comparing ANOVAs 1 and 5 that adding the 23 department categories increases the explanatory power of the model modestly, from .339 to .367. This is another indication of the minor importance of interregional variation in explaining overall inequality.¹⁷

Quantification of Personal and Regional Effects

In order to evaluate the magnitude of various categorical effects with and without standardizing for the effect of other variables, the main-effects model is estimated in equivalent regression form based on dummy variables. As in the ANOVA, the dependent variable is the logarithm of income. As independent variables, the rural-urban and department categories are used along with categorical data on education and age. The regression results are reported in table 6. Regression 1 considers the rural-urban distinction only. The coefficient on the rural dummy variable is $-.981$. This means that measured in logarithms urban workers receive 98% more income than rural workers.¹⁸ Regression 2 includes only information on department of residence, expressed as deviations from Tolima;

¹⁷ When 77 two-way interaction effects are estimated without restrictions, the proportion of the variance explained increases from 0.35 to 0.39. The most significant interactions are between the rural-urban residence distinction and education/age/region, implying that further study should explicate the underlying differences between rural and urban income functions. This is confirmed in research in which the returns to schooling are found to be lower in rural than in urban areas (see Helena Jaramillo, "Determinants of Income Differentials after Migration in Colombia" [paper presented at meeting on Population and Development, Battelle, Seattle, June 1978]).

¹⁸ Recall that the average urban-rural income ratio is three to one. That was based on income as the dependent variable. The 98% differential observed here is based on the logarithm of income as the dependent variable. Equivalently, three to one is the ratio of ordinary arithmetic means and two to one is the ratio of geometric means.

the regression coefficients on the department dummy variables imply that, for example, Bogotá reports incomes 85% more than Tolima and Nariño 74% less. Regression 3 shows that when one holds constant whether the individual resides in an urban or rural area, these department dummy variable coefficients diminish in absolute magnitude; for example, Bogotá becomes +.43 and Nariño -.66, but they do not disappear.

Age and education categories are included without regional variables in regression 4.¹⁹ The coefficients on the education categories show that employees with no education receive incomes 45% less than those with some primary, while employees with secondary education earn nearly twice as much (0.926) in the logarithms and employees with higher education earn nearly three times as much as those with a primary education (1.96). Workers aged 10–19 earn 63% less than workers aged 25–29. Incomes rise with age in the cross section, peaking between 35 and 55, at which age incomes tend to be some 25% higher than for those in the late twenties.

Regression 5 combines employee characteristics with geographic information. Because of covariation between these two sets of influences, the regression coefficients diminish in average absolute magnitude when combined. Comparing regressions 3 and 5 the rural-urban differential decreases from -0.80 to -0.44, a reduction of 45%. The average absolute value of the department dummies decreases by 50%. The age coefficients decrease on average by 9%, and the education coefficients on average by 18%.

These data provide the information needed to quantify the effect of compositional differences in interregional inequality in Colombia. By adjusting for age and education, the gross rural-urban income differential and the average interdepartmental differential are both reduced by about half. In other words, half the interregional differences in incomes in Colombia can be explained simply in terms of age and education. Still, much remains to be accounted for by, on the one hand, other aspects of workers' skills, job experience, and training and, on the other hand, by long-run factor market distortions and short run quasi-rents to workers in specific regional labor markets. However, from the small changes in the regression coefficients on the education and age variables, we conclude that the income differences across regions are not accounted for to a large extent by interregional differentials in rewards to education or experience.

Relative Importance of Interregional Inequality

Relatively how important are the personal and regional effects explaining income inequality in Colombia? The rural-urban distinction alone accounts for 14.5% of the log variance of income in Colombia; information on the 23 departments for 10.7%; and the two together, 18.5%. By comparison, 28.7% of the log variance of income is explained by 10 age and

¹⁹ Once again, these regression coefficients should be understood as the average effect of the explanatory variable expressed as a geometric mean.

Economic Development and Cultural Change

TABLE 6
REGRESSIONS ON THE LOGARITHM OF INCOME BASED ON CATEGORICAL DATA

EXPLANATORY VARIABLE	MALE EMPLOYEES FOR REGRESSION NUMBER:				
	1	2	3	4	5
Education (deviation from primary):					
None	-.453 (20.5)	-.300 (13.7)
Secondary926 (45.5)	.709 (33.3)
Higher	1.96 (45.9)	1.73 (41.5)
Age (deviations from 25-29):					
10-19	-.624 (21.4)	-.575 (20.7)
20-24	-.260 (9.16)	-.242 (8.94)
30-34155 (4.94)	.144 (4.82)
35-44257 (9.12)	.237 (8.83)
45-54252 (7.88)	.218 (7.14)
55+	-.0329 (.87)	-.024 (.67)
Zone, rural-urban (deviations from urban)	-.981 (53.1)	...	-.799 (39.8)	...	-.438 (22.8)
Departments (deviations from Tolima):					
Antioquia367 (7.27)	.271 (5.60)206 (4.77)
Atlántico663 (10.4)	.271 (4.41)125 (2.26)
Bogotá850 (16.7)	.433 (8.72)207 (4.65)
Bolívar355 (4.76)	.119 (1.67)803 (1.31)
Boyacá	-.466 (7.10)	-.357 (5.69)	...	-.413 (7.38)
Caldas206 (3.21)	.192 (3.14)179 (3.26)
Cauca	-.245 (3.08)	.210 (2.76)	...	-.253 (3.74)
César205 (2.27)	.109 (1.26)189 (2.45)
Córdoba	-.073 (1.02)	.026 (.38)071 (1.16)
Cundinamarca	-.103 (1.72)	-.042 (.74)	...	-.095 (1.86)
Chocó	-.462 (3.21)	-.571 (4.15)	...	-.577 (4.71)
Huila	-.113 (1.47)	-.149 (2.04)	...	-.139 (2.14)
La Guajira258 (1.78)	.066 (.47)064 (.52)
Magdalena079 (.97)	.057 (.73)090 (1.30)

TABLE 6 (Continued)

EXPLANATORY VARIABLE	MALE EMPLOYEES FOR REGRESSION NUMBER:				
	1	2	3	4	5
Meta267 (3.01)	.130 (1.54)132 (1.75)
Nariño	-.736 (10.8)	-.658 (10.1)	...	-.712 (12.3)
Norte de Santander	-.186 (2.84)	-.184 (2.94)	...	-.150 (2.69)
Quindío181 (2.20)	.054 (.68)012 (.17)
Risaralda286 (3.86)	.196 (2.77)101 (2.55)
Santander	-.129 (2.19)	-.187 (3.31)	...	-.170 (3.38)
Sucre	-.122 (1.23)	-.174 (1.83)	...	-.088 (1.04)
Valle469 (9.05)	.227 (4.55)174 (3.92)
Intercept	6.88	6.28	6.71	6.41	6.54
R ²1455	.1071	.1852	.2865	.3531
SEE	1.141	1.67	1.115	1.043	.994

NOTE.—*t*-ratios are in parentheses.

education variables. Taken together, the 23 rural-urban and department variables along with the 10 age and education variables explain 35.3% of the log variance.²⁰ A standard *F*-ratio test would suggest the need to include regional effects as part of the explanation for income inequality in Colombia.²¹ However, the 23 regional variables decrease the standard error of estimate by only 5% (from 1.043 to 0.994). This is the other side of the analysis-of-variance findings that the regional variables contributed the least to explaining income inequality.

We conclude that even in a country like Colombia, where interregional income disparities are pronounced, the importance of regional effects in total inequality is surprisingly small. Recognition of place of residence,

²⁰ Half the variance in urban log income using a larger set of explanatory variables including education, experience, city of residence, and parents' education is explained by Gary S. Fields, "Education and Economic Mobility in Colombia" (Center Discussion Paper no. 237, Economic Growth Center, Yale University, September 1975). Comparably high *R*²s have been obtained in Colombia, using somewhat different independent variables, by Bernardo Kugler, "Influencia de la educación en los ingresos del trabajo: El caso Colombiano," *Revista de planeación y desarrollo* 6, no. 2 (1974): 51-74. In Brazil, a notably higher *R*² (nearly .6) is reported, but explanatory variables include sex, on which we stratified the sample, and sector. See Carlos Langoni, "Income Distribution and Economic Development: Brazilian Case" (paper presented at the Econometric Society Third World Congress, Toronto, August 1975). An *R*² of .3 is obtained, also using Brazilian census data, by Fishlow (n. 10 above). In the United States, for white nonfarm males, an *R*² of .3 based on schooling and a quadratic in age is reported by Jacob Mincer (*Schooling, Experience and Earnings* [New York: Colombia University Press, 1974]).

²¹ The marginal *F*-ratio test of any restriction on the main-effects model is not likely to be accepted, given the large size of the working sample (16,680) relative to the number of parameters being fitted (32 in regression 5). See Zvi Griliches, "Wages of Very Young Men," *Journal of Political Economy* 84, no. 4, pt. 2 (August 1976): S69-S86.

Economic Development and Cultural Change

while providing some information on income, may complicate the simple linear model of income determination without adding substantially to its predictive precision. Nonetheless, region may have a greater effect on income and income inequality for some education and age groups than for others. Regional effects within education and age groups are, therefore, investigated next.

Analysis of Variance within Education and Age Categories

It was noted in early tabulations that relative inequality measured either by the coefficient of variation or the variance of the logarithm of income is greater for the least educated (table 2). Other tabulations, not published, indicate lower relative interregional variation in income and steeper age-income profiles for the better educated. Thus we anticipate that as educational attainment increases region will be found to be of diminished importance relative to age in accounting for the long variance of income.

To determine the explanatory importance of regional differences in log incomes at different educational levels, the ANOVA model is reported in table 7 within education classes. The 11 regional groups of departments, the rural-urban distinction, and age are used as explanatory variables. The regional categories contribute more to explaining the variance in log incomes for male employees with primary education or less than for those with secondary or higher education. Age accounts for only 1.2% of the log variance of income for those with no schooling but 19.5% among the higher educated. The department groups directly account for 6% of the log variance with no education, but only 3% for the better educated. The rural-urban distinction is also more important for men with only some primary education than for those with secondary and higher education.²² Thus, as hypothesized, age is more important and region tends to be less important in accounting for inequality as we move to higher levels of education.

Turning to age effects, we might hypothesize that regional differences in incomes would be more notable among older employees, given that the propensity to move declines with age. To explore this question, we performed analyses of variance within age groups, examining the relative contribution of education and place of residence to explained sum of squares (table 8). The results show that less than 6% of the log variance of incomes within age groups is associated with department groupings, declining somewhat to age 30–34 and rising thereafter. The rural-urban distinction accounts for 2%–5% after age 20. Educational categories, on the other hand, explain an increasing share of the log variance within age groups, from 3% at age 10–19 to 18% by age 30–34, and then diminishing to 12% within the oldest age group. Consequently, the relative importance of education vis-à-vis region increases sharply up to age 35, whereupon the

²² With only six employees with higher education in rural areas, the *F*-test for the rural-urban effect is understandably insignificant within the higher education class.

TABLE 7
ANALYSIS OF VARIANCE WITHIN EDUCATION CLASSES, MALE EMPLOYEES

MAIN EFFECTS	df	NONE		PRIMARY		SECONDARY		HIGHER	
		Proportion of Variance Explained	F-Ratio, Marginal	Proportion of Variance Explained	F-Ratio, Marginal	Proportion of Variance Explained	F-Ratio, Marginal	Proportion of Variance Explained	F-Ratio, Marginal
Age.....	6	.012	6.72*	.065	130.*	.187	130.*	.195	26.3*
Department groups.....	10	.061	19.9*	.037	44.1*	.005	2.26	.035	2.81
Rural/urban.....	1	.007	23.3*	.044	524.*	.011	47.0*	.000	.00
Covariance.....	..	.012		.057		.010		-.010	
Total main effects.....	17	.092	17.7*	.203	144.*	.213	52.3*	.220	10.5*
Logarithm of income:									
Mean.....	...		5.92		6.43		7.26		8.44
Variance.....	...		1.17		1.19		1.15		1.26
Sample size.....	...		2,987		9,606		3,302		647

* Statistically significant at .001 level.

TABLE 8
ANALYSIS OF VARIANCE WITHIN AGE GROUPS, MALE EMPLOYEES

VARIABLE	df	10-19		20-24		25-29		30-34		35-44		45-54		55 AND OVER	
		Pro- por- tion of Vari- ance Ex- plained	F- Ratio, Mar- ginal	Pro- por- tion of Vari- ance Ex- plained	F- Ratio, Mar- ginal	Pro- por- tion of Vari- ance Ex- plained	F- Ratio, Mar- ginal	Pro- por- tion of Vari- ance Ex- plained	F- Ratio, Mar- ginal	Pro- por- tion of Vari- ance Ex- plained	F- Ratio, Mar- ginal	Pro- por- tion of Vari- ance Ex- plained	F- Ratio, Mar- ginal	Pro- por- tion of Vari- ance Ex- plained	F- Ratio, Mar- ginal
Education.....	2/3 ^a	.030	32.0*	.053	63.1*	.148	193.0*	.179	203.6*	.162	284.0*	.179	189.4*	.117	63.7*
Department groupings.....	10	.055	17.5*	.021	7.6*	.033	12.9*	.014	4.7*	.020	10.3*	.035	11.1*	.043	7.1*
Rural/urban.....	1	.001	2.1	.018	64.3*	.030	116.4*	.047	160.8*	.044	231.5*	.032	102.5*	.024	39.3*
Covariance.....039082163174177152108	...
Total explained.....124	28.4*	.174	44.7*	.374	104.6*	.415	100.9*	.402	151.5*	.398	90.4*	.294	34.2*
Logarithm of income:															
Mean.....	...	5.86		6.41		6.76		6.86		6.82		6.72		6.32	
Variance.....	...	1.12		1.32		1.37		1.47		1.56		1.50		1.60	
Sample size.....	...	2,821		2,983		2,471		2,005		3,166		1,930		1,166	

^a In the case of ages 10-19 there were no higher-educated male employers, and thus only three education groups were represented in the sample and only 2 degrees of freedom employed.

* Statistically significant at .001 level.

ratio turns down. This is consistent with the age selectivity of migration in equilibrating labor markets, young workers moving at high rates to take advantage of interregional wage disparities but mobility diminishing beyond the midpoint of the life cycle.

IV. Summary and Conclusions

Working with a sample of 860,000 individuals from the 1973 Colombian Census of Population, we have sought to understand the determinants of income and income inequality. The data exhibit noticeable differences in mean income between men and women (25% more for men), between employees and employers (60% more for employers), across education categories (eleven to one between higher educated and none), across regions (four to one between the richest and poorest departments), between urban and rural workers (three to one), and by age (four to one between prime age workers and the very young). These differences arise both in the simple comparisons and in the finer cross-classifications by age, education, region, and urban-rural simultaneously. This led to one conclusion of this study; that interregional inequality is substantial in Colombia even for somewhat "homogeneous" classes of workers.

To interpret variation in income across this large number of groupings, a formal statistical framework is needed. For this purpose, a linear model in the form of analysis of variance (ANOVA) and multiple regression is employed, applied to logarithms of male employee income. The results of the ANOVA supported the hypotheses that education, age, region, and rural-urban variables contribute significantly in accounting for income inequality in Columbia. But their contributions are not equal. By standard statistical conventions, the four-way classification by educational attainment is much the more important, while the single urban-rural dichotomy is next in importance per degree of freedom used. Following these, the seven age categories are generally more significant statistically than the 6, 11, or 23 regional categories. We thus reached a second conclusion: to predict an individual's income, if you could ask only one question, knowing the individual's education would give a more accurate prediction than would either his age, region, or knowing whether he lived in an urban or rural area.

To quantify the various personal and regional effects, both singly and together, regression analysis was then used to compare effects on geometric means. When the several sets of variables were included in a single regression, the gross differentials were altered. In a regression where 24 regional effects were added to personal effects, the standard error of estimate was reduced by only 5%. This provides some justification for ignoring geographic information if a simple income determination model is of interest. Another regression result was that the structures of income differences across education and age groups do not differ greatly from one regional labor market in Colombia to another. Also, in the regressions,

Economic Development and Cultural Change

department of residence was found to account for 11% of income inequality (as measured by the log variance) and department and rural-urban location together for 19%. The regression results yielded a third conclusion: that most (greater than 80%) of income inequality in Colombia is within rather than between regions.

We next turned our attention to the patterns of income inequality for different education and age groups and the correlates of those patterns. A pattern of larger relative dispersion of incomes across regions for the less educated was found both in the tabulations and in the analysis of variance. Across education groups, region's explanatory power was greatest for the lower educational groups, diminishing at higher levels, and age gained in importance as education increased. Across age groups, education became increasingly important up to middle age; the main regional effects were found to be small and exhibited no pronounced trend. These results suggest that if regional labor markets in Colombia are not clearing because of institutional restrictions or inertia of potential migrants, this problem is most severe among the least-educated and among prime age workers. But even for these groups, most inequality is within rather than between regions. In other words, a fourth conclusion is that the weakness of regional factors in explaining income inequality in Colombia continues to be found even when the population is stratified to allow for larger regional effects for some education or age groups than for others.

Our research supports the view that regional effects help to explain income inequality in Colombia, but that these effects are quite limited. Both absolutely and relative to personal variables, region is a rather poor predictor of income. Hence, the importance given in past studies to inter-regional inequality in Colombia may have been somewhat overstated.

This conclusion has policy implications. Pronouncements of public policy in Colombia regularly stress the importance of improving the economic position of the poorer half of the population and reducing income inequality. Since incomes are known to be associated, among other things, with place of residence, some policymakers and advisors have called for allocating development resources to the poorer departments (e.g., Chocó) and to the rural areas to the exclusion of the richer departments (e.g., Bogotá) and of the towns and cities. From our research, we can say that such a screening mechanism is too crude: it would help many nonpoor in low-income areas while automatically excluding large numbers of poor who reside in better-off locations.

However development is defined (whether in terms of aggregate growth, employment generation, inequality reduction, or alleviation of absolute poverty), it is unlikely to be achieved by using average incomes to divide the economy arbitrarily into "deserving" and "undeserving" regions, industries, or sectors. It is wrong to think that the poor are to be found in some segments of Colombia and not in others. To decide how to

raise the income of the poor requires a model of those manipulable factors responsible for the level of personal incomes and hence for the degree of inequality. We have made a start by considering age, education, and geographic information as factors conditioning income. But to proceed to policy prescriptions requires more elaborate and realistic models of personal income formation than we have presented here. A promising direction for future research in accounting for income inequality, particularly in the rural sector among the self-employed, is to develop procedures to take account of the distribution of ownership of the nonlabor factors of production (land and capital).